PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A23J 1/00, C12N 15/63, 15/00, 1/21, C12P 21/04, A61K 38/00, C07K 1/00, C07H 21/04

A1

(11) International Publication Number: **WO 99/57992**

(43) International Publication Date: 18 November 1999 (18.11.99)

(21) International Application Number: PCT/US99/10662

(22) International Filing Date: 14 May 1999 (14.05.99)

(30) Priority Data:
09/078,687 14 May 1998 (14.05.98) US

(71) Applicant: CLONTECH LABORATORIES, INC. [US/US]; 1020 East Meadow Drive, Palo Alto, CA 94303 (US).

(72) Inventors: TCHAGA, Grigoriy; 35073 Dorchester Court, Newark, CA 94560 (US). JOKHADZE, George, G.; 360 Chiquita Avenue #9, Mountain View, CA 94041 (US).

(74) Agent: ADLER, Benjamin, A.; McGregor & Adler, 8011 Candle Lane, Houston, TX 77071 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: COMPOSITIONS AND METHODS FOR PROTEIN PURIFICATION BASED ON A METAL ION AFFINITY SITE

(57) Abstract

The present invention describes introducing metal ion affinity sequences into recombinant proteins to allow for purification and/or immobilization of these proteins. The present invention provides a fusion protein, comprising: a protein of interest fused at its amino–terminus or carboxy–terminus to at least one affinity peptide, said fusion protein having a formula R1–$(HX_n)_m$–R2, wherein R1 or R2 is said protein of interest, n = 1–8, m = 2–30, and wherein if n = 1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

COMPOSITIONS AND METHODS FOR PROTEIN PURIFICATION BASED ON A METAL ION AFFINITY SITE

BACKGROUND OF THE INVENTION

Cross Reference to Related Applications

This application claims benefit of non-provisional application US Serial number 09/078.687 filed May 14, 1998.

Field of the Invention

This invention relates generally to the field of protein chemistry. Specifically, the present invention relates to protein purification using a metal ion affinity site.

Background of the Invention:

Development of protocols for the isolation and purification of proteins is often a long and costly process. Such protocols usually contain multiple steps, where some of the steps have recoveries as low as 50%. Further, due to variation between protein molecules, a purification protocol developed and effective for the purification of one protein is not necessarily useful for the purification of another. In fact, in most cases considerable

adaptations must be made to a purification protocol to accommodate the various physical and chemical characteristics of different proteins.

The ability to prepare hybrid genes by genetic engineering technology has opened up new possibilities for the purification of proteins. For example, one can link a DNA sequence of a protein of interest to a nucleic acid sequence which codes for a peptide which has a high binding affinity for a specific ligand. The fusion protein product resulting from expression of this DNA has attributes of both the protein of interest and the high affinity peptide. To purify or immobilize the engineered fusion protein, the ligand commonly is linked to a support, and the unpurified, engineered protein is then exposed to the ligand/support composite and allowed to bind.

There are numerous advantages of using a high affinity fusion protein. For example, the use of an affinity peptide ensures that no part of the native protein of interest is involved in adsorption—the binding between the fusion protein and the ligand. At the same time, extremely high selectivity in the adsorption process is achieved.

Immobilized Metal Ion Affinity Chromatography (IMAC) is one of the most frequently used techniques for purification of fusion proteins containing affinity sites for metal ions. Proper choice of immobilized metal ion, loading conditions and elution conditions can result in protein purification of up to about 95-98% in a single chromatographic step. Moreover, recovery generally is higher than 85%. In addition to the advantages discussed above, incorporation of a proteolytic, chemical, or enzymatic cleavage site into the composite DNA,

between the affinity peptide and the sequence of the protein of interest, provides a means for cleaving the affinity peptide from the protein of interest to yield the native protein of interest in highly purified form.

The following publications are representative of the art: Itakura, et al., *Science* 198:1056-63 (1977); Germino, et al., *PNAS USA* 80:6848-52 (1983); Nilsson et al., *Nucleic Acid Res.* 13:1151-62 (1985); Smith et al., *Gene* 32:321-27 (1984); Dobeli, et al., U.S. Pat. No. 5,284,933; and Dobeli, et al., U.S. Pat. No. 5,310,663.

The prior art is deficieint in improved compositions and methods for affinity immobilization and purification of proteins. This invention fulfills this long-felt need in the art.

SUMMARY OF THE INVENTION

The present invention relates to compositions and methods for protein purification involving the use of novel, genetically-engineered fusion proteins. These fusion proteins are engineered to allow for immobilization and purification via the high affinity interaction of an affinity peptide of a fusion protein with a ligand. The affinity peptide is a histidine-rich polypeptide sequence with a general sequence: $(HX_n)_m$, wherein H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine. The affinity peptide

is linked to the proteins of interest R1 and R2 to yield a fusion protein with formula R1-$(HX_n)_m$-R2. In a preferred embodiment of the invention, n=1-4 and m=2-10. In a more preferred embodiment of the invention, n=1-4 and m=3-6. In a specific embodiment of the invention, a fusion protein having the sequence SLKDHLIHNVHKEEHAHAHNKISVVGVGAVGM (SEQ ID No:6) is provided. In another embodiment of the present aspect of the invention, at least one protease cleavage site is inserted between the sequence of the protein of interest and the sequence of the affinity peptide.

In another aspect of the invention, there is provided a DNA sequence coding for a fusion protein comprising a protein of interest fused at its amino-terminus or carboxy-terminus to at least one affinity peptide, where the fusion protein has the general formula R1-$(HX_n)_m$-R2, wherein R1 or R2 is the protein of interest, H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine. In a specific embodiment of this aspect of the invention, there is provided a DNA sequence which codes for a protein where the fusion protein has the sequence SLKDHLIHNVHKEEHAHAHNKISVVGVGAVGM (SEQ ID No:6).

In various embodiments of this aspect of the invention, there is provided a recombinant vector comprising an expression vector and a DNA sequence coding for a fusion protein comprising a protein of interest fused at its amino-terminus or carboxy-terminus to at least one affinity peptide as described above, wherein the recombinant vector is capable of directing

expression of the DNA sequence in a suitable host organism. The present invention also provides a host organism containing a recombinant expression vector comprising a DNA sequence coding for a fusion protein comprising a protein of interest fused at its amino-terminus or carboxy-terminus to at least one affinity peptide as described above, wherein the organism is capable of expressing said DNA sequence.

In yet an additional aspect of this invention, there is provided a method for purifying the novel fusion proteins of the present invention, comprising the steps of: contacting a protein sample containing the fusion protein in a mixture with other proteins with a metal chelate resin under conditions where the fusion protein binds to the resin to produce a resin-fusion protein complex; washing the resin-fusion protein complex with a buffer to remove the other, unbound proteins; and eluting the bound fusion protein from the washed resin-fusion protein complex. One embodiment of this method includes inserting at least one protease cleavage site between the protein of interest and the affinity peptide, and cleaving the protein of interest from the affinity peptide after purification using the metal chelate resin.

Other and further aspects, embodiments, features and advantages of the present invention will be apparent from the following description of the invention.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic representation of the pGFPuv/HAT vector. This vector contains one embodiment of the

affinity peptide of the present invention fused to the N-terminus of a UV mutant of Green Fluorescent Protein. Only unique restriction sites are noted.

Figure 2 is a schematic representation of the vector pUC19/HS) containing part of the affinity peptide (AP) at the N-terminus of Enterokinase (EK) cleavage site followed by multiple cloning site (MCS). Only restriction sites are denoted.

Figure 3 is a schematic representation of the restriction maps of a vector with three frame shifts containing part of the affinity peptide and enterokinase cleavage site that is used for expression of recombinant proteins.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to compositions and methods for purification of novel, genetically-engineered fusion proteins. Immobilization and purification is achieved via a high-affinity interaction of an affinity peptide portion of the fusion protein with a ligand. The high affinity peptide is a histidine-rich polypeptide with a general sequence $(HX_n)_m$ which is linked to proteins of interest R1 or R2; wherein H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine. This high affinity polypeptide is incorporated by fusion to the N- or C-terminal sequence of a protein of interest and is used for high selectivity purification of the fusion protein.

The affinity of the high affinity peptide is for immobilized metal ions. The strength of binding between the high affinity peptide and an appropriate metal ion is very high; thus, isolation of the fusion proteins is very selective. However, association between the peptide and the ligand is also reversible. Once the fusion protein has been allowed to associate or adsorb with the metal ion ligand, the protein can be disassociated or eluted from the metal ion/adsorbent by addition of competitive ligand such as imidazole, or by decreasing the pH, which leads to protonation of the nitrogen in the imidazole ring of the histidine side chain and release of the adsorbed protein. Because of this reversibility, the protein is recovered in a purified, unbound form. Further, regeneration and reuse of the metal ion/adsorbent or support multiple times--even more than 100 times--is possible.

An additional feature of the protein purification and immobilization techniques based on the principles of the present invention is the high probability that the purified and regenerated protein of interest will retain full biological activity and specificity. This is because the affinity peptide is involved in the immobilization/binding process where the portion of the fusion protein that contains the protein of interest is not.

Incorporation of a proteolytic site between the high affinity peptide and the sequence of the protein of interest provides the means to regenerate the protein of interest from the fusion protein. Regeneration is achieved by limited proteolysis of the fusion protein and a second chromatography step in which the proteolytic product is passed through an immobilized metal ion affinity column. Indeed, one can utilize the same column as was used to immobilize and purify the fusion protein. In the second

chromatography step following proteolysis, the cleaved, regenerated protein of interest passes through the column without immobilization or adsorption, whereas the high affinity peptide is adsorbed on the column.

One embodiment of the present invention features incorporation of nucleic acid sequences which code for secretion signals into the DNA sequence that codes for the fusion protein. Such secretion signals cause the fusion protein to be secreted into the media after synthesis in a host cell. Since a considerable amount of total cellular protein remains in the cell, secretion improves dramatically the isolation and purification of the fusion protein by eliminating the need for cell disruption, protein extraction, and/or removal of unwanted cellular components and nucleic acid.

As used herein, the terms "affinity peptide" or "high affinity peptide" refer to a histidine-rich polypeptide with a general sequence $(HX_n)_m$ which is linked to proteins of interest R1 or R2; wherein H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine.

As used herein, the term "protein of interest" shall refer to any protein to which the affinity peptide is fused for the purpose of purification or immobilization.

As used herein, the term "fusion protein" shall refer to the protein hybrid containing the affinity peptide and the protein of interest or any amino acid sequence of interest. The fusion protein has the general formula R1-$(HX_n)_m$-R2; wherein R1 or R2 is

the protein of interest, H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine.

As used herein, the terms "secretion sequence" or "secretion signal sequence" shall refer to an amino acid signal sequence which leads to the transport of a protein containing the signal sequence outside the cell membrane. In the present case, a fusion protein of the present invention may contain such a secretion sequence to enhance and simplify purification. Representative examples of secretion signal sequences are well known to those having ordinary skill in this art.

As used herein, the term "proteolytic site" shall refer to any amino acid sequence recognized by any proteolytic enzyme. In the present case, a fusion protein of the present invention may contain such a proteolytic site between the protein of interest and the affinity peptide and/or other amino acid sequences so that the protein of interest may be separated easily from these heterologous amino acid sequences.

As used herein, the term "metal ion" refers to any metal ion for which the affinity peptide has affinity and that can be used for purification or immobilization of a fusion protein. Such metal ions include Ni^{+2}, Co^{+2}, Zn^{+2}, Cu^{+2}, Ac^{+3} and Fe^{+3}.

As used herein, the terms "adsorbent" or "solid support" shall refer to a chromatography or immobilization medium used to immobilize a metal ion.

As used herein, the term "regeneration", in the context of the fusion protein, shall refer to the process of separating or

eliminating the affinity peptide and other heterologous amino acid sequences from the fusion protein to render the protein of interest after purification.

So that the matter in which the above-recited features, advantages, and objects of the invention become clear and can be understood in detail, particular descriptions of the invention may be had by reference to particular embodiments described in the Examples below; however, the following description and examples are given for the purpose of illustrating various, specific embodiments of the invention, and are not meant to limit the scope of the invention in any fashion.

EXAMPLE 1

Extraction of lactate dehydrogenase from chicken breast muscle

A naturally-occurring peptide sequence from the N-terminus of lactate dehydrogenase (LDH) from chicken muscle (*Gallus gallus*) was used for initial experiments. The protein includes a stretch of approximately 30 amino acids which has a sequence consistent with the general formula of the fusion protein of the present invention (SLKDHLIHNVHKEEHAHAHNKISVVGVGAVGM (SEQ ID No:6)). Further, LDH has the feature that the enzyme itself can be assayed easily for activity. Thus, the naturally-occurring chicken muscle LDH served as a "fusion protein" for these experiments in the sense that it contained both a high affinity peptide and a protein of interest.

Extraction of chicken breast muscle LDH was performed by cutting 15 g of frozen chicken muscle, free of blood vessels, into small pieces and transferring the material to a commercial blender along with 150 mL of extraction buffer (50 mM sodium phosphate, 1 mM EDTA, 1 mM magnesium acetate pH 7.5, 1 mM 2-mercaptoethanol (0.2 L) stored for at least 30 minutes at 4°C). The mixture was homogenized twice at 4°C for 30 seconds, with a 10-minute pause between the bursts. After the second homogenization, the mixture was transferred to centrifuge tubes and centrifuged at 4°C and 10,000 x g for 30 minutes. The clear supernatant was collected and used as a starting sample for the purification of lactate dehydrogenase.

EXAMPLE 2

Purification of lactate dehydrogenase on Ni(II)-chelating sepharose FF

Lactate dehydrogenase was purified by IMAC in the following manner: approximately 5 mL of Chelating Sepharose FF (Amersham, Pharmacia) was transferred to a vacuum bottle, diluted with an equal volume of deionized water and degassed under vacuum for 10 minutes. The gel suspension was poured into a column (10x1 cm. i. d.) trapped on the bottom with a degassed adapter and left to settle. The column was then filled to the top with degassed deionized water, and a top adapter was gently pushed down the column bed until there was no space between the top surface of the gel and the adapter. The column

was washed with 3 column volumes of deionized water at a flow rate of 0.5 mL per min.

The chicken muscle extract (14 mL) was equilibrated by gel filtration on Sephadex G-25 columns with equilibration buffer (20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 0.06 M imidazole pH 7.0 (1 L)). The IMAC column was then charged with Ni(II) ions using 20 mL of a 0.02M $Ni(NO_3)_2$ solution. The excess metal was washed from the column with deionized water at a flow rate of 0.5 mL per minute and the column was then equilibrated with 5-10 volumes of equilibration buffer (20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 0.06 M imidazole pH 7.0 (1 L)).

The IMAC column was prepared by loading the equilibrated extract on to the IMAC column at a flow rate of 0.5 mL per min. Fractions of 1 mL were collected. The column was washed with equilibration buffer until a baseline was reached (absorbance of the fractions at 280 nm was less than 2 mAU higher than the absorbance of the equilibration buffer). The adsorbed material was eluted with elution buffer (20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 0.3 M imidazole pH 7.0 (0.2 L)) and absorbance at 280 nm was determined on a spectrophotometer. Protein content of each fraction was determined as described in M. Bradford, *Analytical Biochemistry*, 72 (1976) 248, and lactate dehydrogenase activity was determined as described in F. Kubowitz and P. Ott, *Biochem. Z*, 314 (1943) 94. Results indicated that more than 95% of the lactate dehydrogenase activity was recovered in the elution fractions.

EXAMPLE 3

Characterization of lactate dehydrogenase binding

Further experiments were performed both with native LDH, a tetramer of about 140 kD, and a subunit of the enzyme, obtained after warming of the crude chicken muscle extract to 45°C for 10 minutes. Both the tetramer and the subunit were allowed to associate with the immobilized Ni support, and both forms of LDH were retained. This result demonstrates that the retention of the LDH enzyme on immobilized Ni is not peculiar to the tetrameric form of the peptide; that is, binding does not require "cooperation" between subunits. Instead, the single subunit of the enzyme also had affinity for the nickel ion, and this affinity was demonstrated to be virtually identical to the affinity shown for the tetramer. Both the native protein and the subunit were adsorbed in buffer with an imidazole concentration up to 60 mmol and both were eluted completely at a concentration of 300 mmol imidazole.

To ascertain that it is the polyhistidine portion of the LDH that provides affinity for the nickel ion, the tetramic form of the LDH enzyme was subjected to CNBr cleavage to produce a mixture of peptides. This mixture of peptides was applied to a Ni-IDA column with metal ion capacity of 32 mmol per mL gel. Loading conditions were the same as those used for the purification of the enzyme from the crude extract, described above. The adsorbed material was eluted with 300 mmol imidazole, and subjected to RPC chromatography. The chromatographic peak containing about 80% of the adsorbed

material was then subjected to amino acid analysis. The results obtained demonstrate that this peak corresponds to the N-terminal peptide from LDH and that this peptide that contains the polyhistidine sequence. In addition, the fact that the peptide retained its binding affinity even after treatment with CNBr in presence of 70% TFA is proof that the binding is not due to a rigid secondary conformation structure.

EXAMPLE 4

Purification of lactate dehydrogenase on Co2+-TALON agarose

Extraction of chicken breast muscle LDH was performed as in Example 1, and equilibrated by gel filtration on Sephadex G-25 columns with equilibration buffer (20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 0.06 M imidazole pH 7.0 (1 L)).

The IMAC column was prepared in the following manner: Approximately 2.75 mL of Co2+-TALON Superflow 6 (Amersham, Pharmacia) was transferred to a vacuum bottle, diluted with the same volume of deionized water and degassed under vacuum for 10 minutes. The gel suspension was poured into a column (3x1 cm. i.d.) trapped on the bottom with a degassed adapter and left to settle. The column was filled to the top with degassed deionized water, and a top adapter was gently pushed down toward the column bed until there was no space between the top surface of the gel and the adapter. The column was washed with 3 column volumes of deionized water at a flow rate of 0.5 mL per min.

Purification of the fusion protein on Co2+-TALON Superflow 6 was carried out by first equilibrating the IMAC column with 5 to 10 column volumes of the equilibration buffer. The sample was then loaded on the IMAC column at a flow rate of 1.0 mL per min, and 1 mL fractions were collected. The column was washed with the equilibration buffer until a baseline was reached (absorbance of the fractions at 280 nm as less than 2 mAU higher than the absorbance of the equilibration buffer).

The adsorbed material was eluted with elution buffer (20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 0.3 M imidazole pH 7.0 (0.2 L)) and absorbance at 280 nm was determined on a spectrophotometer. Protein content of each fraction was determined as described in M. Bradford, *Analytical Biochemistry*, 72 (1976) 248, and lactate dehydrogenase activity was determined as described in F. Kubowitz and P. Ott, *Biochem. Z.*, 314 (1943) 94. As in Example 2, more than 95% of the lactate dehydrogenase activity was recovered in the elution fractions.

EXAMPLE 5

Isolation and purification of fusion protein consisting of affinity peptide and Green Fluorescent Protein UV Mutant (GFPuv)

An affinity peptide/GFP fusion protein was isolated from *E. coli* cells which had been transformed with the pGFPuv.HS vector. Cell paste (0.39 g) was transferred to pre-cooled mortar, 1.2 g of alumina was added, and the mixture was ground for 2 minutes. Extraction buffer (5 mL, stored at 4°C) was added, and,

after additional grinding for 2 minutes, the mixture was transferred into four eppendorph tubes. The suspension was added to the eppendorph tubes and centrifuged for 12 minutes at 12,000 rpm (11,750 x g). The clear supernatant (approximately 6 mL) was used as a starting sample for IMAC.

The extraction and chromatography equilibration buffers consisted of 20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 5 mM imidazole pH 7.0 (1 L). The elution buffer for IMAC consisted of 20 mM sodium phosphate buffer containing 1.0 M sodium chloride and 150 mM imidazole pH 7.0 (0.2 L).

The IMAC was carried out in the following manner: Approximately 2.75 mL of Co2+-TALON Superflow 6 (Amersham, Pharmacia) was transferred to a vacuum bottle, diluted with the same volume of deionized water and degassed under vacuum for 10 minutes. The gel suspension was poured into a column (3x1 cm. i.d.) trapped on the bottom with a degassed adapter and left to settle. The column was filled to the top with degassed deionized water, and a top adapter was gently pushed down toward the column bed until there was no space between the top surface of the gel and the adapter. The column was washed with 3 column volumes of deionized water at a flow rate of 0.5 mL per min.

Purification of the fusion protein on Co2+-TALON Superflow 6 was carried out by first equilibrating the IMAC column with 5 to 10 column volumes of the equilibration buffer. The sample was then loaded on the IMAC column at a flow rate of 1.0 mL per min, and 1 mL fractions were collected. The column was washed with the equilibration buffer until a baseline was

reached (absorbance of the fractions at 280 nm as less than 2 mAU higher than the absorbance of the equilibration buffer). The adsorbed material was then eluted with elution buffer.

Absorbance of each fraction at 280 nm was determined on a spectrophotometer; and protein content of each fraction was determined. Fluorescence of each fraction was determined on a microplate reader, and the purity of the fusion protein was determined also by SDS-electrophoresis. More than 85% of the fusion protein was recovered in the fractions obtained. Part of the cDNA sequence, and the amino acid sequence encoded by this cDNA sequence, of a vector containing the affinity peptide at the N-terminus of Green Fluoresecent Protein-UV mutant (GFPuv) is shown in SEQ ID No. 1 and SEQ ID No. 2, respectively. The full cDNA sequence of a vector containing the construct of the affinity peptide at the N-terminus of GFPuv is shown in SEQ ID No. 3. The full cDNA sequence of a vector containing part of the affinity peptide at the N-terminus of the enterokinase cleavage site and the amino acid sequence encoded by this cDNA corresponding to the start of translation site, the affinity peptide and the multiple cloning site are shown in SEQ ID Nos. 4 and 5.

EXAMPLE 6

Construction of fusion proteins

A DNA sequence corresponding to the affinity peptide of the present invention is fused to the DNA coding sequence of a protein of interest. The polynucleotide sequence for the affinity peptide is fused most generally at or close to the DNA sequence

coding for the N- or C-terminal amino acid of the protein of interest. This results in a DNA sequence which codes for a fusion protein comprising the affinity peptide and the protein of interest.

In addition, a polynucleotide sequence that codes for a protein proteolytic site is incorporated into the fusion protein DNA sequence between the sequence for the affinity peptide and the sequence of the protein of interest. This type of DNA construct results in a fusion protein product having a proteolytic site. This site allows for the eventual regeneration of the protein of interest from the fusion protein by limited proteolysis and a second chromatography step. The second chromatography step, in which the product of the proteolysis is loaded onto an immobilized metal ion affinity column, results in the separation of the protein of interest from the affinity peptide.

An additional embodiment of the present invention provides a DNA sequence coding for a polypeptide "secretion signal" introduced into the DNA that codes for the fusion protein. This secretion signal, when expressed, causes the fusion protein to be secreted into the culture media after the fusion protein is synthesized in the cell. Since a large number of cellular proteins are not transported out of the cell, isolation and purification of the fusion protein is enhanced as the requirements for cell disruption, extraction and removal of unwanted cell components are eliminated.

The present invention is directed to a fusion protein of general formula R1-$(HX_n)_m$-R2 comprising: a protein of interest R1 or R2 fused at its amino terminus or carboxy terminus to at least one affinity peptide, said affinity peptide having a formula $(HX_n)_m$, wherein H is histidine, X is an amino acid other than histidine, n=

1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine. Preferably, n=1-4 and m=3-10. In one preferred embodiment, n=1-4 and m=3-6. Preferably, if n=1 for more than two adjacent units of HX, only one X is asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine. In one preferred embodiment, the fusion protein has a sequence SEQ ID No. 1. The fusion protein may contain at least one protease cleavage site between said protein of interest and said affinity peptide. Preferably, the affinity peptide has affinity for metal ions. A representative metal ion is a nickel ion. The fusion protein may further comprise a secretion signal sequence.

The present invention is also directed to a DNA sequence coding for a fusion protein of general formula R1-$(HX_n)_m$-R2 comprising: a protein of interest R1 or R2 fused at its amino terminus or carboxy terminus to at least one affinity peptide, said affinity peptide having a formula $(HX_n)_m$, wherein H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine. Preferably, the fusion protein has the sequence shown in SEQ ID No:6.

The present invention is also directed to a recombinant vector comprising a DNA sequence disclosed herein, wherein said recombinant vector is capable of directing expression of said DNA sequence for said fusion protein in a suitable host organism. The present invention is also directed to a

host organism containing a recombinant vector disclosed herein, wherein said organism is capable of expressing said fusion protein.

The present invention is also directed to a method for purifying the fusion protein disclosed herein, comprising the steps of:

contacting a protein sample containing said fusion protein in a mixture with other proteins with a metal chelate resin under conditions where said fusion protein binds to said resin to produce a resin-fusion protein complex; washing said resin-fusion protein complex with a buffer to remove said other, unbound proteins; and eluting said bound fusion protein from the washed resin-fusion protein complex; wherein said eluted fusion protein is purified. This method may further comprise the step of cleaving said protein of interest from said affinity peptide. Moreover, this method further comprises the step of separating said cleaved protein of interest from said affinity peptide using a. metal chelate resin under conditions where said affinity peptide binds to said metal of said resin and said protein of interest does not.

Any patents or publications mentioned in this specification are indicative of the levels of those skilled in the art to which the invention pertains. These patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

One skilled in the art will readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The present examples along with the methods,

procedures, treatments, molecules, and specific compounds described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention as defined by the scope of the claims.

WHAT IS CLAIMED IS:

1. A fusion protein comprising: a protein of interest fused at its amino terminus or carboxy terminus to at least one affinity peptide, said fusion protein having a formula R1-$(HX_n)_m$-R2, wherein R1 or R2 is said protein of interest, H is histidine, X is an amino acid other than histidine, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine.

2. The fusion protein of claim 1, wherein n=1-4.

3. The fusion protein of claim 1, wherein m=3-10.

4. The fusion protein of claim 1, wherein n=1-4 and m=3-6.

5. The fusion protein of claim 1, wherein if n=1 for more than two adjacent units of HX, only one X is asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine.

6. The fusion protein of claim 2, wherein said fusion protein has a sequence SEQ ID No. 6.

7. The fusion protein of claim 1, wherein said fusion protein contains at least one protease cleavage site between said protein of interest and said affinity peptide.

8. The fusion protein of claim 1, wherein said affinity peptide has affinity for metal ions.

9. The fusion protein of claim 8, wherein said metal ions are nickel ions.

10. The fusion protein of claim 1, wherein said fusion protein further comprises a secretion signal sequence.

11. A DNA sequence coding for a fusion protein comprising a protein of interest fused at its amino- or carboxy-terminus to at least one affinity peptide, said fusion protein having the formula R1-$(HX_n)_m$-R2, wherein R1 or R2 is said protein of interest, n= 1-8, m= 2-30, and wherein if n=1 for more than two adjacent units of HX, at least one X must be asparagine, phenylalanine, tryptophan, tyrosine, lysine, methionine, arginine, glutamine, or cysteine.

12. The DNA sequence of claim 11, wherein said fusion protein has the sequence shown in SEQ ID No:1.

13. A recombinant vector comprising a DNA sequence of claim 11, wherein said recombinant vector is capable of directing expression of said DNA sequence for said fusion protein in a suitable host organism.

14. A host organism containing a recombinant vector of claim 13, wherein said organism is capable of expressing said fusion protein.

15. A method for purifying the fusion protein of claim 1, comprising the steps of:

contacting a protein sample containing said fusion protein in a mixture with other proteins with a metal chelate resin under conditions where said fusion protein binds to said resin to produce a resin-fusion protein complex;

washing said resin-fusion protein complex with a buffer to remove said other, unbound proteins; and

eluting said bound fusion protein from the washed resin-fusion protein complex; wherein said eluted fusion protein is purified.

16. The method of claim 15, further comprising the step of cleaving said protein of interest from said affinity peptide.

17. The method of claim 16, further comprising the step of separating said cleaved protein of interest from said affinity peptide using a. metal chelate resin under conditions where said affinity peptide binds to said metal of said resin and said protein of interest does not.

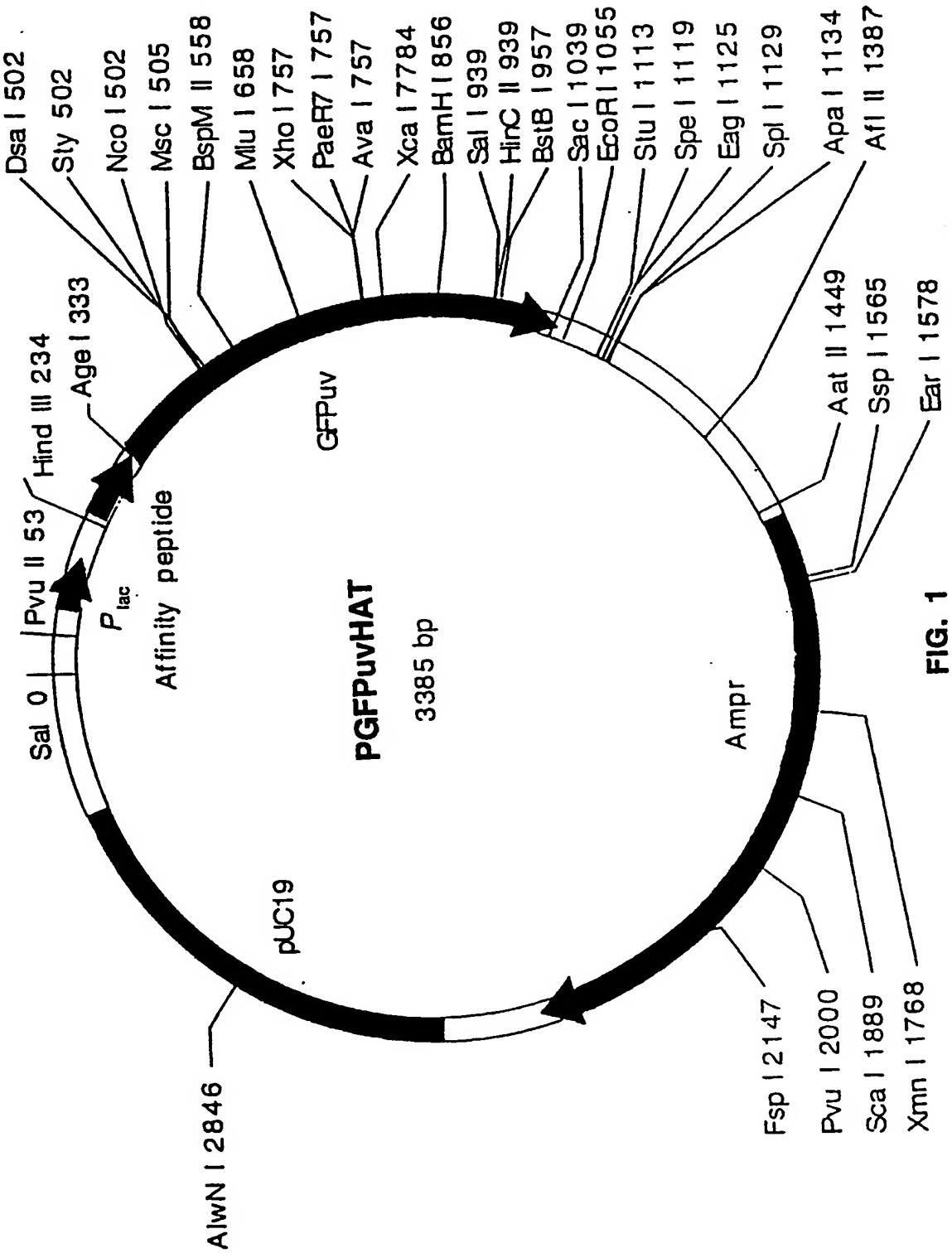
PGFPuvHAT
3385 bp
GFPuv
Ampr
pUC19
Affinity peptide
P_{lac}
Sal 0
Pvu II 53
Hind III 234
Age I 333
Dsa I 502
Sty 502
Nco I 502
Msc I 505
BspM II 558
Mlu I 658
Xho I 757
PaeR7 I 757
Ava I 757
Xca I 7784
BamH I 856
Sal I 939
HinC II 939
BstB I 957
Sac I 1039
EcoR I 1055
Stu I 1113
Spe I 1119
Eag I 1125
Spl I 1129
Apa I 1134
Afl II 1387
Aat II 1449
Ssp I 1565
Ear I 1578
AlwN I 2846
Fsp I 2147
Pvu I 2000
Sca I 1889
Xmn I 1768

FIG. 1

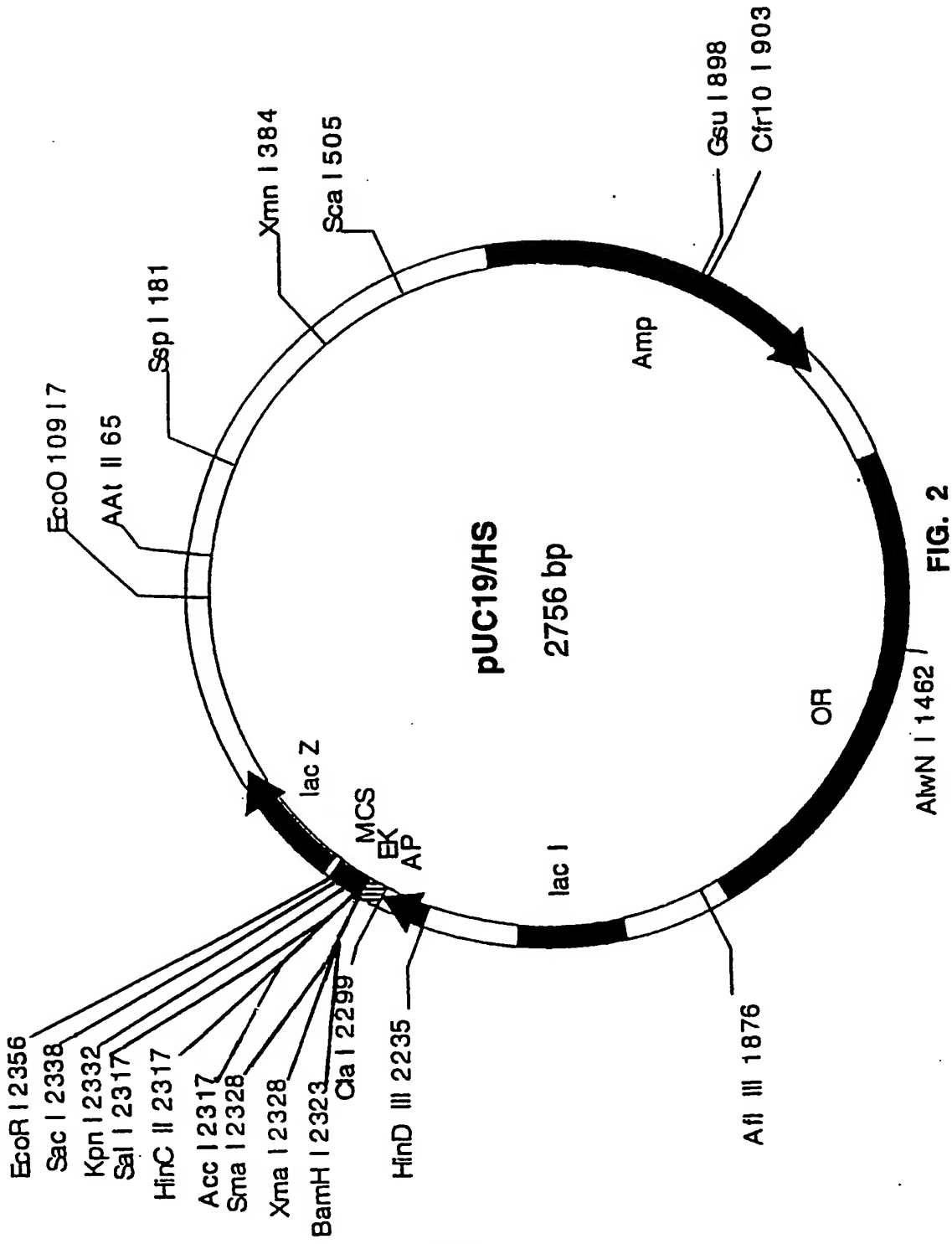
pUC19/HS
2756 bp
Amp
OR
lac I
lac Z
MCS
EK
AP
EcoR I 2356
Sac I 2338
Kpn I 2332
Sal I 2317
HinC II 2317
Acc I 2317
Sma I 2328
Xma I 2328
BamH I 2323
Cla I 2299
HinD III 2235
Afl III 1876
AlwN I 1462
Cfr10 I 903
Gsu I 898
Sca I 505
Xmn I 384
Ssp I 181
AAt II 65
EcoO 109 I 7

FIG. 2

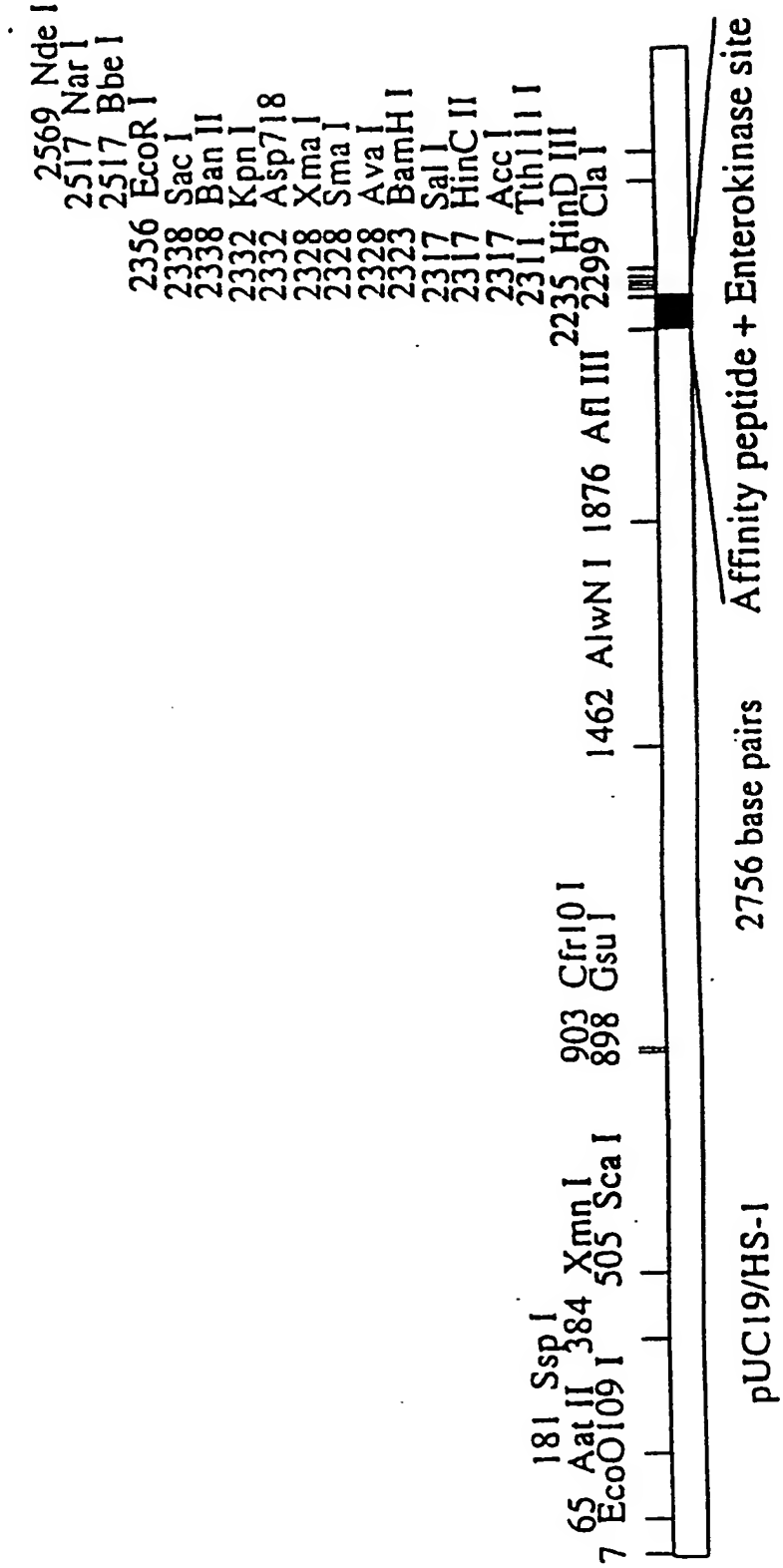
2569 Nde I
2517 Nar I
2517 Bbe I
2356 EcoR I
2338 Sac I
2338 Ban II
2332 Kpn I
2332 Asp718
2328 Xma I
2328 Sma I
2328 Ava I
2323 BamH I
2317 Sal I
2317 HinC II
2317 Acc I
2311 Tth111 I
2235 HinD III
2299 Cla I
1876 Afl III
1462 AlwN I
903 Cfr10 I
898 Gsu I
505 Sca I
384 Xmn I
181 Ssp I
65 Aat II
7 EcoO109 I
pUC19/HS-1
2756 base pairs
Affinity peptide + Enterokinase site

FIG. 3-1

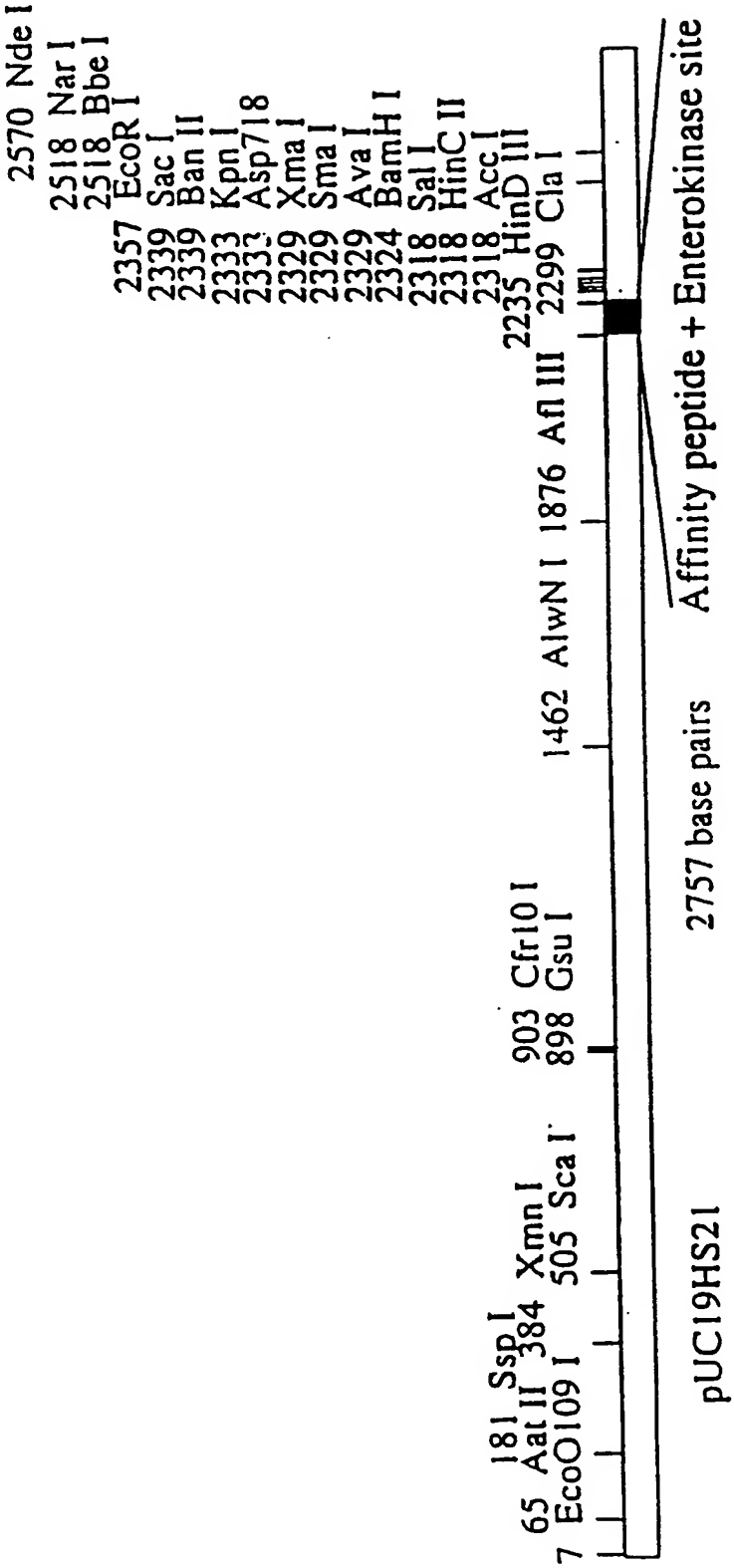
2570 Nde I
2518 Nar I
2518 Bbe I
2357 EcoR I
2339 Sac I
2339 Ban II
2333 Kpn I
2333 Asp718
2329 Xma I
2329 Sma I
2329 Ava I
2324 BamH I
2318 Sal I
2318 HinC II
2318 Acc I
2235 HinD III
2299 Cla I
1876 Afl III
1462 AlwN I
903 Cfr10 I
898 Gsu I
505 Sca I
384 Xmn I
181 Ssp I
65 Aat II
7 EcoO109 I
pUC19HS21
2757 base pairs
Affinity peptide + Enterokinase site

FIG. 3-2

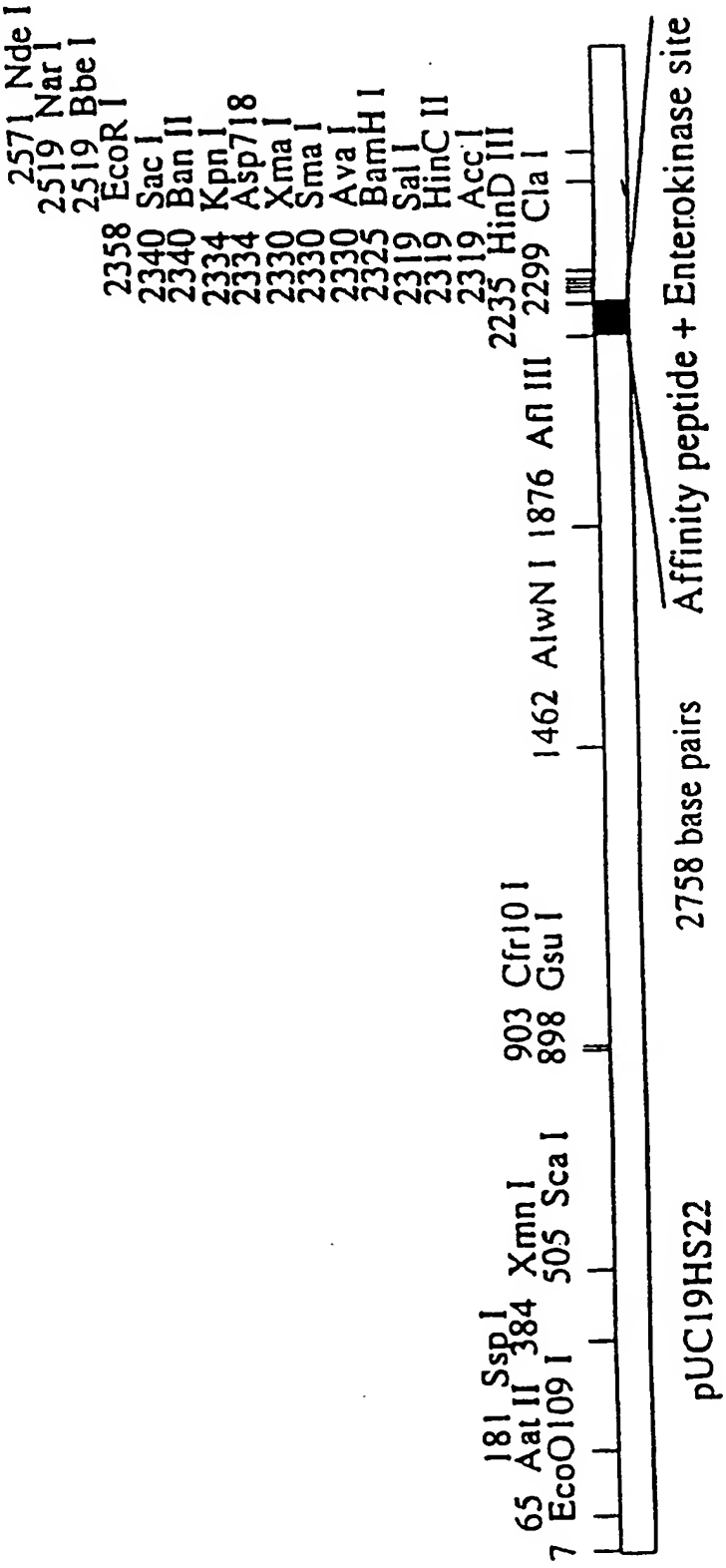
2571 Nde I
2519 Nar I
2519 Bbe I
2358 EcoR I
2340 Sac I
2340 Ban II
2334 Kpn I
2334 Asp718
2330 Xma I
2330 Sma I
2330 Ava I
2325 BamH I
2319 Sal I
2319 HinC II
2319 Acc I
2235 HinD III
2299 Cla I
1876 Afl III
1462 AlwN I
903 Cfr10 I
898 Gsu I
505 Sca I
384 Xmn I
181 Ssp I
65 Aat II
7 EcoO109 I
pUC19HS22
2758 base pairs
Affinity peptide + Enterokinase site

FIG. 3-3

SEQUENCE LISTING

<110> Tchaga, Grigoriy
Jokhadze, George G.

<120> Compositions and Methods for Protein Purification Based on a Novel Metal Ion Affinity Site

<130> D6094PCT

<140>

<141> 1999-05-14

<150> US 09/078,687

<151> 1998-05-14

<160> 6

<210> 1

<211> 840

<212> DNA

<213> artificial sequence

<220>

<223> Partial cDNA sequence of a vector containing the affinity peptide at the N-terminus of Green Fluorescent Protein-UV mutant (GFPuv)

<400> 1

atgaccatga ttacgccaag cttgtctctc aaggatcatc tcatccacaa 50
tgtccacaaa gaggagcacg ctcatgccca caacaagatc agcgtggttg 100
gtgtgggtgc agttggaccg gtaagtaaag gagaagaact tttcactgga 150
gttgtcccaa ttcttgttga attagatggt gatgttaatg ggcacaaatt 200
ttctgtcagt ggagagggtg aaggtgatgc aacatacgga aaacttaccc 250
ttaaatttat ttgcactact ggaaaactac ctgttccatg gccaacactt 300
gtcactactt tctcttatgg tgttcaatgc ttttcccgtt atccggatca 350
tatgaaacgg catgactttt tcaagagtgc catgcccgaa ggttatgtac 400
aggaacgcac tatatctttc aaagatgacg ggaactacaa gacgcgtgct 450
gaagtcaagt ttgaaggtga taccttgtt aatcgtatcg agttaaaagg 500
tattgatttt aaagaagatg gaaacattct cggacacaaa ctcgagtaca 550
actataactc acacaatgta tacatcacgg cagacaaaca aaagaatgga 600
atcaaagcta acttcaaaat tcgccacaac attgaagatg gatccgttca 650
actagcagac cattatcaac aaaatactcc aattggcgat ggccctgtcc 700
ttttaccaga caaccattac ctgtcgacac aatctgccct ttcgaaagat 750
cccaacgaaa agcgtgacca catggtcctt cttgagtttg taactgctgc 800
tgggattaca catggcatgg atgagctcta caaataatga 840

<210> 2
<211> 278
<212> PRT
<213> artificial sequence
<220>
<223> Amino acid sequence encoded by partial cDNA sequence of a vector containing the affinity peptide at the N-terminus of Green Fluorescent Protein-UV mutant (GFPur
<400> 2

```
Met Thr Met Ile Thr Pro Ser Leu Ser Leu Lys Asp His Leu Ile
 5                  10                  15
His Asn Val His Lys Glu Glu His Ala His Ala His Asn Lys Ile
            20                  25                  30
Ser Val Val Gly Val Gly Ala Val Gly Pro Val Ser Lys Gly Glu
        35                  40                  45
Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly
    50                  55                  60
Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly
65                  70                  75
Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr
                80                  85                  90
Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe Ser
            95                  100                 105
Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Arg
        110                 115                 120
His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
    125                 130                 135
Arg Thr Ile Ser Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala
140                 145                 150
Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu
155                 160                 165
lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys
            170                 175                 180
Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Thr Ala Asp
        185                 190                 195
Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Ile Arg His Asn
    200                 205                 210
```

Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn
215 220 225
Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
230 235 240
Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg
245 250 255
Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr
260 265 270
His Gly Met Asp Glu Leu Tyr Lys
275

<210> 3
<211> 3384
<212> DNA
<213> artificial sequence
<220>
<223> cDNA sequence of a vector containing the construct of the affinity peptide at the N-terminus of GFPuv
<400> 3

agcgcccaat acgcaaaccg cctctccccg cgcgttggcc gattcattaa 50
tgcagctggc acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa 100
cgcaattaat gtgagttagc tcactcatta ggcaccccag gctttacact 150
ttatgcttcc ggctcgtatg ttgtgtggaa ttgtgagcgg ataacaattt 200
cacacaggaa acagctatga ccatgattac gccaagcttg tctctcaagg 250
atcatctcat ccacaatgtc cacaaagagg agcacgctca tgcccacaac 300
aagatcagcg tggttggtgt gggtgcagtt ggaccggtaa gtaaaggaga 350
agaacttttc actggagttg tcccaattct tgttgaatta gatggtgatg 400
ttaatgggca caaattttct gtcagtggag agggtgaagg tgatgcaaca 450
tacggaaaac ttacccttaa atttatttgc actactggaa aactacctgt 500
tccatggcca acacttgtca ctactttctc ttatggtgtt caatgctttt 550
cccgttatcc ggatcatatg aaacggcatg actttttcaa gagtgccatg 600
cccgaaggtt atgtacagga acgcactata tctttcaaag atgacgggaa 650
ctacaagacg cgtgctgaag tcaagtttga aggtgatacc cttgttaatc 700
gtatcgagtt aaaaggtatt gattttaaag aagatggaaa cattctcgga 750
cacaaactcg agtacaacta taactcacac aatgtataca tcacggcaga 800
caaacaaaag aatggaatca aagctaactt caaaattcgc cacaacattg 850
aagatggatc cgttcaacta gcagaccatt atcaacaaaa tactccaatt 900
ggcgatggcc ctgtcctttt accagacaac cattacctgt cgacacaatc 950

tgcccttcg aaagatccca acgaaaagcg tgaccacatg gtccttcttg 1000
agtttgtaac tgctgctggg attacacatg gcatggatga gctctacaaa 1050
taatgaattc caactgagcg ccggtcgcta ccattaccaa cttgtctggt 1100
gtcaaaaata taggcctac tagtcggccg tacgggccct ttcgtctcgc 1150
gcgtttcggt gatgacggtg aaaacctctg acacatgcag ctcccggaga 1200
cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag 1250
ggcgcgtcag cgggtgttgg cgggtgtcgg ggctggctta actatgcggc 1300
atcagagcag attgtactga gagtgcacca tatgcggtgt gaaataccgc 1350
acagatgcgt aaggagaaaa taccgcatca ggcggcctta agggcctcgt 1400
gatacgccta tttttatagg ttaatgtcat gataataatg gtttcttaga 1450
cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc tatttgttta 1500
tttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg 1550
ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt 1600
tccgtgtcgc ccttattccc ttttttgcgg cattttgcct tcctgttttt 1650
gctcacccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg 1700
tgcacgagtg ggttacatcg aactggatct caacagcggt aagatccttg 1750
agagttttcg ccccgaagaa cgttttccaa tgatgagcac ttttaaagtt 1800
ctgctatgtg gcgcggtatt atcccgtatt gacgccgggc aagagcaact 1850
cggtcgccgc atacactatt ctcagaatga cttggttgag tactcaccag 1900
tcacagaaaa gcatcttacg gatggcatga cagtaagaga attatgcagt 1950
gctgccataa ccatgagtga taacactgcg gccaacttac ttctgacaac 2000
gatcggagga ccgaaggagc taaccgcttt tttgcacaac atgggggatc 2050
atgtaactcg ccttgatcgt tgggaaccgg agctgaatga agccatacca 2100
aacgacgagc gtgacaccac gatgcctgta gcaatggcaa caacgttgcg 2150
caaactatta actggcgaac tacttactct agcttcccgg caacaattaa 2200
tagactggat ggaggcggat aaagttgcag gaccacttct gcgctcggcc 2250
cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg 2300
gtctcgcggt atcattgcag cactggggcc agatggtaag ccctcccgta 2350
tcgtagttat ctacacgacg gggagtcagg caactatgga tgaacgaaat 2400
agacagatcg ctgagatagg tgcctcactg attaagcatt ggtaactgtc 2450
agaccaagtt tactcatata tactttagat tgatttaaaa cttcattttt 2500
aatttaaaag gatctaggtg aagatccttt ttgataatct catgaccaaa 2550
atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa 2600
gatcaaagga tcttcttgag atccttttt tctgcgcgta atctgctgct 2650
tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa 2700
gagctaccaa ctctttttcc gaaggtaact ggcttcagca gagcgcagat 2750
accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga 2800
actctgtagc accgcctaca tacctcgctc tgctaatcct gttaccagtg 2850

gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg 2900
atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca 2950
cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag 3000
cgtgagctat gagaaagcgc cacgcttccc gaagggagaa aggcggacag 3050
gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc 3100
caggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc 3150
tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg 3200
gaaaaacgcc agcaacgcgg cctttttacg gttcctggcc ttttgctggc 3250
cttttgctca catgttcttt cctgcgttat cccctgattc tgtggataac 3300
cgtattaccg cctttgagtg agctgatacc gctcgccgca gccgaacgac 3350
cgagcgcagc gagtcagtga gcgaggaagc ggaa 3384

<210> 4
<211> 2754
<212> DNA
<213> artificial sequence
<220>
<223> cDNA sequence of a vector containing part of the affinity peptide at the N-terminus of enterokinase cleavage site
<400> 4

gacgaaaggg cctcgtgata cgcctatttt tataggttaa tgtcatgata 50
ataatggttt cttagacgtc aggtggcact tttcggggaa atgtgcgcgg 100
aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca 150
tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt 200
atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt 250
ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300
ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac 350
agcggtaaga tccttgagag ttttcgcccc gaagaacgtt ttccaatgat 400
gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtattgacg 450
ccgggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg 500
gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 550
aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca 600
acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg 650
cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct 700
gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctgtagcaa 750
tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct 800

tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 850
acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 900
gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat 950
ggtaagccct cccgtatcgt agttatctac acgacgggga gtcaggcaac 1000
tatggatgaa cgaaatagac agatcgctga gataggtgcc tcactgatta 1050
agcattggta actgtcagac caagtttact catatatact ttagattgat 1100
ttaaaacttc atttttaatt taaaaggatc taggtgaaga tccttttga 1150
taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 1200
cagaccccgt agaaaagatc aaaggatctt cttgagatcc tttttttctg 1250
cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt 1300
ttgtttgccg gatcaagagc taccaactct ttttccgaag gtaactggct 1350
tcagcagagc gcagatacca aatactgtcc ttctagtgta gccgtagtta 1400
ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 1450
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 1500
ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga 1550
acgggggggtt cgtgcacaca gcccagcttg gagcgaacga cctacaccga 1600
actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag 1650
ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag 1700
cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt 1750
cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 1800
gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc 1850
ctggcctttt gctggccttt tgctcacatg ttctttcctg cgttatcccc 1900
tgattctgtg gataaccgta ttaccgcctt tgagtgagct gataccgctc 1950
gccgcagccg aacgaccgag cgcagcgagt cagtgagcga ggaagcggaa 2000
gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta 2050
atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca 2100
acgcaattaa tgtgagttag ctcactcatt aggcacccca ggctttacac 2150
tttatgcttc cggctcgtat gttgtgtgga attgtgagcg gataacaatt 2200
tcacacagga aacagctatg accatgatta cgccaagctt gaaggatcat 2250
ctcatccaca atgtccacaa agaggagcac gctcatgccc acaacaagat 2300
cgatgacgat gacaaagtcg acggatcccc gggtaccgag ctcgtaatta 2350
gctgagaatt cactggccgt cgttttacaa cgtcgtgact gggaaaaccc 2400
tggcgttacc caacttaatc gccttgcagc acatcccccc ttcgccagct 2450
ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca acagttgcgc 2500
agcctgaatg gcgaatggcg cctgatgcgg tattttctcc ttacgcatct 2550
gtgcggtatt tcacaccgca tatggtgcac tctcagtaca atctgctctg 2600
atgccgcata gttaagccag ccccgacacc cgccaacacc cgctgacgcg 2650
ccctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac 2700

cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac 2750
gcgc 2754

<210> 5
<211> 45
<212> PRT
<213> artificial sequence
<220>
<223> Amino acid sequence encoded by cDNA corresponding to the start of translation site, the affinity peptide and the multiple cloning site
<400> 5

Met Thr Met Ile Thr Pro Ser Leu Lys Asp His Leu Ile His Asn
5 10 15
Val His Lys Glu Glu His Ala His Ala His Asn Lys Ile Asp Asp
20 25 30
Asp Asp Lys Val Asp Gly Ser Pro Gly Thr Glu Leu Val Ile Ser
35 40 45

<210> 6
<211> 32
<212> PRT
<213> artificial sequence
<220>
<223> Amino acid sequence of fusion protein where n=1-4 for the affinity peptide
<400> 6

Ser Leu Lys Asp His Leu Ile His Asn Val His Lys Glu Glu His
5 10 15
Ala His Ala His Asn Lys Ile Ser Val Val Gly Val Gly Ala Val
20 25 30
Gly Met

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/10662

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :435/69.7, 252.3, 320.1; 530/ 811, 413, 350, 324; 536/23.4

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.7, 252.3, 320.1; 530/ 811, 413, 350, 324; 536/23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN, search terms: fusion, metal chelate affinity chromatography, histidine, his, lactate dehydrogenase

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,594,115 A (SHARMA) 14 January 1997 (14/01/97), see entire document, especially column 6, lines 36-37, column 9, lines 1-10, columns 10-13, columns 19-20.	1-17
X	HIROTA et al. Nucleotide and deduced amino acid sequences of chicken lactate dehydrogenase-A. Nuc. Acid. Res. 1990, Vol. 18, No. 21, page 6432, especially column 2, amino acid positions 2-33.	1-9, 11
A	HEMDAN et al. Development of immobilized metal affinity chromatography II. Interaction of amino acids with immobilized nickel iminoacetate. J. Chromatography. 1985, Vol. 323, pages 255-264.	1-17

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
26 JULY 1999	08 SEP 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer HOLLY SCHNIZER Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/10662

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4,569,794 A (SMITH et al.) 11 February 1986 (11/02/86), see entire document, especially column 1, lines 21-25, column 4, lines 38-68, Table I, and examples 4-7.	1-5, 7-9, 11, 15-17
Y	SMITH et al. Chelating peptide-immobilized metal ion affinity chromatography. J. Biol. Chem. 25 May 1988, Vol. 263, No. 15, pages 7211-7215, especially page 7212, columns 1 and 2, and page 7214, column 2.	1-4, 8, 9, 11, 15-17
Y	STADER et al. 'Engineering *Escherichia coli* to secrete heterologous gene products.' In: Methods in Enzymology. Edited by Abelson et al. San Diego: Academic Press, Inc. 1990, Vol. 185, pages 166-193, especially page 167.	10
Y	LJUNGQUIST et al. Immobilization and affinity purification of recombinant proteins using histidine peptide fusions. Eur. J. Biochem. 1989, Vol. 186, pages 563-568, especially page 563, column 1, line 12-15, page 564-565.	1-5, 8-9, 11, 15-17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/10662

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A23J 1/00; C12N 15/63, 15/00, 1/21; C12P 21/04; A61K 38/00; C07K 1/00; C07H 21/04